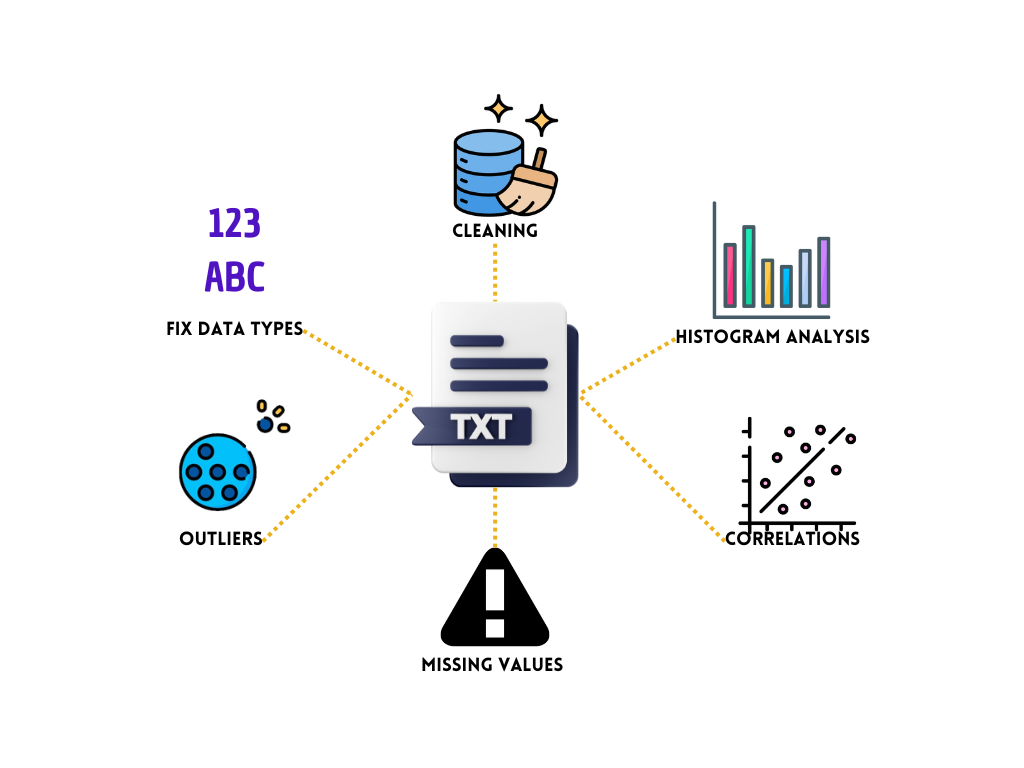
This project aimed to perform a descriptive analysis of a bank's past-due portfolio database. It sought to identify starting points for creating a report to improve the management of overdue accounts through Data Science practices. Data cleaning activities, analysis of distributions, correlations, and identification of outliers were performed.

**\*\*Here are the questions I was interested in answering\*\***

1) What are the most relevant variables for my analysis?

2) What data segmentation should I use initially: city, type of loan, balance due, days past due?

3) Which variables am I interested in using for correlation analysis?

4) What is the final information needed to start the data science stage?

**\*\*I took the following steps to create my analysis\*\***

1) Review all variables and leave only the relevant ones for my analysis.

2) Review and convert number and text formats.

3) Perform a frequency analysis to identify patterns.

4) Use a Pareto chart to identify the city with the highest debt representation.

5) Analyze outliers to avoid errors in future analysis.

6) Analyze correlations between variables.

**\*\*Here are my takeaways\*\***

1) The analysis was reduced from 138 variables to 24 relevant variables.

2) It was identified that there is missing information in the database.

3) The most representative city in terms of days past due and accumulated capital was determined.

4) The team was informed of the final information needed for the predictive analysis, including data on correlations, type of loan, city, principal balance due, and sociodemographic profile.